nature genetics

Genome-wide association study of prostate cancer identifies a second risk locus at 8q24

Meredith Yeager[1,2], Nick Orr[3], Richard B Hayes[2], Kevin B Jacobs[4], Peter Kraft[5], Sholom Wacholder[2], Mark J Minichiello[6], Paul Fearnhead[7], Kai Yu[2], Nilanjan Chatterjee[2], Zhaoming Wang[1,2], Robert Welch[1,2], Brian J Staats[1,2], Eugenia E Calle[8], Heather Spencer Feigelson[8], Michael J Thun[8], Carmen Rodriguez[8], Demetrius Albanes[2], Jarmo Virtamo[9], Stephanie Weinstein[2], Fredrick R Schumacher[5], Edward Giovannucci[10], Walter C Willett[10], Geraldine Cancel-Tassin[11], Olivier Cussenot[11], Antoine Valeri[11], Gerald L Andriole[12], Edward P Gelmann[13], Margaret Tucker[2], Daniela S Gerhard[14], Joseph F Fraumeni Jr[2], Robert Hoover[2], David J Hunter[2,5], Stephen J Chanock[2,3] & Gilles Thomas[2]

Recently, common variants on human chromosome 8q24 were found to be associated with prostate cancer risk. While conducting a genome-wide association study in the Cancer Genetic Markers of Susceptibility project with 550,000 SNPs in a nested case-control study (1,172 cases and 1,157 controls of European origin), we identified a new association at 8q24 with an independent effect on prostate cancer susceptibility. The most significant signal is 70 kb centromeric to the previously reported SNP, rs1447295, but shows little evidence of linkage disequilibrium with it. A combined analysis with four additional studies (total: 4,296 cases and 4,299 controls) confirms association with prostate cancer for rs6983267 in the centromeric locus ($P = 9.42 \times 10^{-13}$; heterozygote odds ratio (OR): 1.26, 95% confidence interval (c.i.): 1.13–1.41; homozygote OR: 1.58, 95% c.i.: 1.40–1.78). Each SNP remained significant in a joint analysis after adjusting for the other (rs1447295 $P = 1.41 \times 10^{-11}$; rs6983267 $P = 6.62 \times 10^{-10}$). These observations, combined with compelling evidence for a recombination hotspot between the two markers, indicate the presence of at least two independent loci within 8q24 that contribute to prostate cancer in men of European ancestry. We estimate that the population attributable risk of the new locus, marked by rs6983267, is higher than the locus marked by rs1447295 (21% versus 9%).

In developed countries, prostate cancer is the most common non-cutaneous malignancy in men, yet a positive family history of prostate cancer and ethnic background are the only established risk factors[1–3]. In the USA, men of African descent are at greater risk than those of European descent[2]. Two independent studies previously demonstrated a single nucleotide polymorphism (SNP) in 8q24, rs1447295, is associated with prostate cancer risk[4,5]. In one study, a stronger association was observed in African Americans[4], while the other study reported a stronger association with aggressive prostate cancer[5]. A third larger study, nested in seven USA and European cohorts and including more than 7,000 prostate cancer cases and 8,000 matched controls, reported an association between rs1447295 and increased risk for prostate cancer in Caucasian men, regardless of age at diagnosis ($P = 4.00 \times 10^{-19}$)[6].

We conducted a genome-wide association study (GWAS) of 550,000 SNPs in 1,172 affected individuals (484 with nonaggressive prostate cancer, Gleason <7 and stage A/B; 688 aggressive prostate cancer, Gleason ≥ 7 and/or stage C/D) and 1,157 controls using an incidence density sampling strategy in the Prostate, Lung, Colon and Ovarian (PLCO) Trial[7,8] (see **Supplementary Methods** online and the Cancer Genetic Markers Susceptibility website). The GWAS confirmed the association for rs1447295, located at physical position 128554220 in NCBI genome build 36 ($P = 9.75 \times 10^{-5}$; heterozygote OR: 1.42, 95% c.i.: 1.16–1.73; homozygote OR: 2.78, 95% c.i.: 1.32–5.86; **Table 1**).

[1]SAIC-Frederick, National Cancer Institute (NCI)-Frederick Cancer Research and Development Center, Frederick, Maryland 21702, USA. [2]Division of Cancer Epidemiology and Genetics and [3]Pediatric Oncology Branch, Center for Cancer Research, NCI, US National Institutes of Health (NIH), Department of Health and Human Services (DHHS), Bethesda, Maryland 20892, USA. [4]Bioinformed Consulting Services, Gaithersburg, Maryland 20877, USA. [5]Program in Molecular and Genetic Epidemiology, Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [6]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SA, UK. [7]Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK. [8]Department of Epidemiology and Surveillance Research, American Cancer Society, Atlanta, Georgia 30329, USA. [9]Department of Health Promotion and Chronic Disease Prevention, National Public Health Institute, Helsinki, FIN-00300, Finland. [10]Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [11]Centre de Recherche pour les Pathologies Prostatiques (CeRePP), Hôpital Tenon, Assistance Publique-Hôpitaux de Paris, 75970 Paris, France. [12]Division of Urologic Surgery, Washington University School of Medicine, St. Louis, Missouri 63108, USA. [13]Division of Hematology and Oncology, Columbia University, New York, New York 10032, USA. [14]Office of Cancer Genomics, NCI, NIH, DHHS, Bethesda, Maryland 20892, USA. Correspondence should be addressed to S.J.C. (chanocks@mail.nih.gov).

Received 16 January 2006; accepted 7 March 2007; published online 1 April 2007; doi:10.1038/ng2022

Table 1 Results of the single-SNP analysis of rs6983267, rs1447295 and rs7837688 (per study and combined)

		Number of subjects	Number of controls	Number of affected individuals	Risk allele frequency, affected individuals	Risk allele frequency, controls	*P*	OR (GT)	95% c.i.	OR (GG)	95% c.i.
rs6983267	PLCO	2,329	1,157	1,172	0.55	0.49	2.43 × 10-05	1.40	1.14–1.73	1.73	1.37–2.19
	ACS	2,301	1,151	1,150	0.55	0.5	3.16 × 10-03	1.27	1.03–1.56	1.49	1.18–1.89
	ATBC	1,790	896	894	0.57	0.51	1.89 × 10-03	1.26	0.99–1.60	1.66	1.28–2.16
	FPCC	914	459	455	0.56	0.51	1.17 × 10-01	1.13	0.81–1.59	1.45	1.00–2.10
	HPFS	1,261	636	625	0.57	0.51	9.54 × 10-03	1.15	0.87–1.53	1.58	1.15–2.16
	ALL	8,595	4,299	4,296	0.56	0.5	9.42 × 10-13	1.26	1.13–1.41	1.58	1.40–1.78
								OR (CA)		OR (AA)	
rs1447295	PLCO	2,329	1,157	1,172	0.14	0.10	9.75 × 10-05	1.42	1.16–1.73	2.78	1.32–5.86
	ACS	2,301	1,151	1,150	0.12	0.08	2.26 × 10-05	1.53	1.23–1.90	2.82	1.30–6.14
	ATBC	1,790	896	894	0.21	0.17	2.88 × 10-02	1.24	1.01–1.52	1.64	0.98–2.72
	FPCC	914	459	455	0.12	0.07	4.35 × 10-03	1.71	1.20–2.43	3.11	0.62–15.71
	HPFS	1,261	636	625	0.13	0.09	2.74 × 10-03	1.56	1.18–2.06	2.51	0.77–8.20
	ALL	8,595	4,299	4,296	0.15	0.11	1.53 × 10-14	1.43	1.29–1.59	2.23	1.58–3.14
								OR (GT)		OR (TT)	
rs7837688	PLCO	2,327	1,156	1,171	0.14	0.10	6.52 × 10-06	1.52	1.24–1.87	2.78	1.36–5.67
	ACS	2,296	1,150	1,146	0.12	0.08	6.82 × 10-06	1.65	1.33–2.06	2.23	1.00–5.01
	ATBC	1,783	894	889	0.19	0.17	2.85 × 10-01	1.14	0.92–1.40	1.38	0.80–2.38
	FPCC	909	456	453	0.12	0.07	3.45 × 10-03	1.73	1.21–2.48	3.34	0.66–16.82
	HPFS	1,255	634	621	0.14	0.09	1.28 × 10-03	1.60	1.21–2.12	2.56	0.78–8.39
	ALL	8,570	4,290	4,280	0.14	0.10	1.85 × 10-14	1.46	1.32–1.63	2.03	1.43–2.89

Analysis adjusted for age in five-year intervals and study. See **Supplementary Table 1** for genotype distribution information and **Supplementary Table 2** for complete results of each study using unconstrained and multiplicative models. ACS: American Cancer Society Prevention Study II; ATBC: Alpha-Tocopherol, Beta-Carotene Prevention Study; FPCC: CeRePP French Prostate Case-Control Study; HPFS = Health Professionals Follow-up Study; PLCO = Prostate, Lung, Colon, Ovarian Trial.

It also identified four SNPs (rs6983267, rs7837328, rs7014346 and rs12334695) significantly associated with prostate cancer in a second region of low correlation (**Fig. 1**). These SNPs reside in a block[9] with strong linkage disequilibrium bounded by markers rs10505476 (128477298) and rs6470517 (128529586) (**Fig. 1**). We investigated one of the most significant associations (rs6983267) in a combined analysis with four additional replication studies totaling 3,124 affected individuals and 3,142 controls (the American Cancer Society Cancer Prevention Study II[10], 1,150 affected individuals and 1,151 controls; the Health Professionals Follow-up Study[11], 625 affected individuals and 636 controls; the CeRePP French Prostate Case-Control Study[12], 455 affected individuals and 459 controls; and the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study[13], 896 affected individuals and 894 controls) (**Table 1**, **Supplementary Table 1** and **Supplementary Table 2** online). Our results show that rs6983267, which has an overall population frequency of 50% in northern Europeans for the 'at-risk' G allele, replicated in all four studies ($P = 1.63 \times 10^{-8}$), thus providing strong evidence for its contribution to prostate cancer risk.

An additional SNP, rs7837688, from the same bin ($r^2 = 0.81$ with rs1447295 in the PLCO) seemed to be more significant in the GWAS, but in the replication studies, its significance and magnitude of effect were comparable to rs1447295 (**Table 1**) overall. Because rs1447295 has been established as the benchmark, based on replication of the two initial studies[6], we conducted our subsequent analyses with rs1447295.

Markers rs1447295 and rs6983257 are physically close, but the two association signals correspond to two independent loci (**Fig. 1**). Analysis of PLCO controls between SNPs rs10505476 (128,477,298) and rs7837688 (128,608,542) using the SequenceLDhot program[14] gives strong evidence for a hotspot of recombination between rs7841264 and rs1447293 (between 128535996 and 128541502) ($P = 5 \times 10^{-5}$). This corresponds to an inferred location of a recombination hotspot in the HapMap data (data release 21)[15,16]. We estimate that 90% of the meiotic recombination events occurring in the 130-kb region bound by rs10505476 and rs7837688 take place in the 5.5-kb region between rs7841264 and rs1447293. Specifically, the population scaled recombination rate[14] within this hotspot is estimated to be 260 (95% c.i.: 100–540), whereas the recombination rate across the remaining region is approximately 30. As a rough guide to these estimates, an effective population size of 10,000 yields a genetic distance of 0.65 cM within the hotspot, and of 0.075 cM for the remainder of the region. The population-scaled recombination rate governs the amount of linkage disequilibrium that would be expected across the hotspot within our population. Such a large value suggests that there will be almost no linkage disequilibrium across the hotspot.

To further explore this region of 8q24, we performed analyses using inferred ancestral recombination graphs (ARGs)[17] (**Fig. 2** and **Supplementary Note** online). We inferred 100 ARGs for 197 SNPs in the region and tested the genealogies at each SNP position for evidence of association using a four–degree of freedom test (one control phenotype, two case phenotypes and three genotypes), with significance calculated using a maximum of 10^5 permutations. The analysis identified two close, but distinct, regions of strong association that straddle the recombination hotspot marked by rs7841264. Within each region, pairwise comparison of the genealogies at positions of strong association uncovered high correlation. The genealogies were not correlated when the loci were located in different regions. Thus, the ARG procedure detected in each region a single and specific genetic event. The ARGs were used to estimate the frequency of the inferred predisposing alleles in the two regions. In the centromeric region at rs6983267, the frequency of the inferred predisposing allele

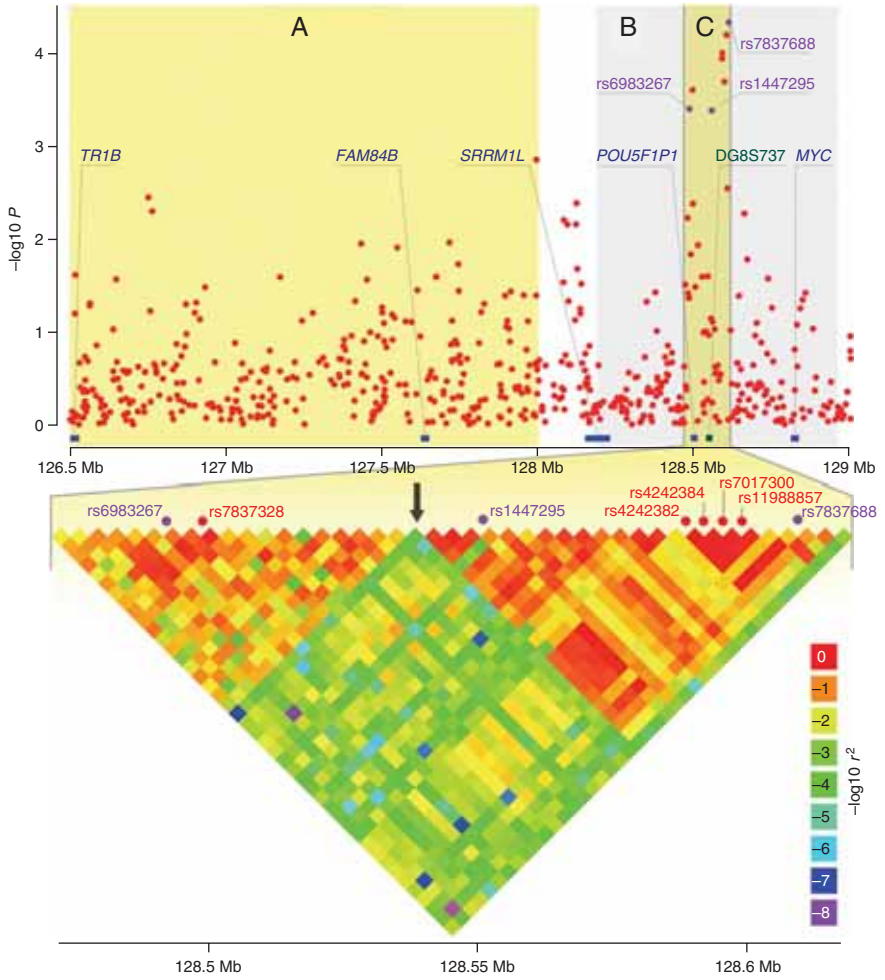

Figure 1 Association analysis of SNPs across a region of 8q24. The upper panel shows *P* values for association testing drawn from a genome-wide association study of prostate cancer in the PLCO cohort across a region of 8q24 bounded by rs4559257 and rs7387606 (chromosome 8: 126501167–128998553). The analysis is based on the genome-wide association study using incidence density sampling with a score test (4 d.f.) adjusted for population stratification (**Supplementary Methods**). Note that rs6983267 is close to a putative pseudogene, *POU5F1P1*. Shaded region 'A' corresponds to the admixture peak reported in ref. 4 and spans chromosome 8, approximately (126500000–128000000). Shaded region 'B' was analyzed by ancestral recombination graph[17] (see **Figs. 2** and **3**). Shaded region 'C' includes a segment containing the most significant *P* values bounded by rs1562871 and rs4407842. Lower panel shows an enlarged view of the region bounded by rs1562871 and rs4407842 (chromosome 8: 128470954–128619305). We estimated the squared correlation coefficient (r^2) for each pairwise comparison of SNPs in this region using a modified version of TagZilla (see Methods). Negative $\log_{10} r^2$ was plotted using Aabel (see Methods) (for example, 0 indicates evidence for strong LD, whereas –8 corresponds to minimal LD). Purple dots represent the locations of the three loci evaluated in this study, rs6983267, rs1447295 and rs7837688 (r^2 for the latter two is 0.81 in the PLCO study); red dots represent the other SNPs that showed evidence of association in the GWAS (rs7837328, rs4242382, rs4242384, rs7017300, and rs11988857). Black arrow indicates the site of recombination discussed in the text.

in controls was 0.46 ± 0.13. In the telomeric region, the frequency was lower (0.12 ± 0.17) at rs1447295 and was between 0.10 and 0.11 (±0.08) for the other five locations (rs4242382, rs4242384, rs7017300, rs11988857 and rs7837688). These differences in estimated allele frequency corroborate the existence of two distinct functional polymorphisms on either side of the recombination hotspot.

In order to identify the haplotypes harboring the deleterious allele of the centromeric region, we phased genotypes from 20 SNPs to determine the most likely pair of haplotypes present in each individual[18,19] (**Fig. 3**). We then used these haplotypes to generate 100 ARGs[17]. At each SNP location, the position of the mutation on inferred genealogies that best explains the disease status partitions the haplotypes into two groups: those predicted to harbor the protective allele versus those predicted to harbor the 'at-risk' allele. At rs6983267, six haplotypes (with frequencies greater than 0.1%) defined by 11 contiguous SNPs were predicted to harbor the protective allele more than 80% of the time and all included the T allele of rs6983267. In the 19 remaining haplotypes, the deleterious allele was predicted in more than 95 of 100 genealogies. All harbored the 'at risk' G allele of rs6983267 (**Fig. 3**). The diversity of the two haplotype groups suggests that the protective allele of the centromeric region is more recent and/or was positively selected. A similar analysis performed with 27 SNPs of the telomeric region identified two haplotypes, defined by 24 contiguous SNPs, carrying the susceptibility alleles, whereas 25 haplotypes carried the protective allele (**Fig. 3**). Taken together, these observations suggest that the telomeric mutational event might have occurred more recently than the centromeric mutation. Furthermore, our data suggest that the telomeric mutation generated a deleterious allele, in contrast to the centromeric mutation.

To determine possible interaction between the two independent SNPs, we investigated seven logistic models for the joint effect of both rs6983267 and rs1447295 on prostate cancer risk, comparing all cases with controls adjusted for study and/or study center and age in 5-year intervals (**Table 2** and **Supplementary Table 3** online). The association between each SNP and prostate cancer risk remained significant after adjusting for the other SNP (in the unconstrained model, rs6983267 $P = 6.62 \times 10^{-10}$ (adjusted for rs1447295 in row 25 of **Supplementary Table 2**); rs1447295 $P = 1.41 \times 10^{-11}$ (adjusted for rs6983267 in row 11 of **Supplementary Table 2**)). No compelling evidence differentiated the unconstrained model (which estimated that odds ratios for the eight nonreferent genotypes varied freely) from the simple multiplicative allelic risk model (likelihood ratio test comparing nested models, $P > 0.7$). The estimated OR under the multiplicative model is 3.17 (95% c.i.: 2.55–3.94) for individuals who are homozygous for the risk alleles at both loci relative to the referent category: namely, double homozygotes for the protective alleles ($P = 9.18 \times 10^{-22}$).

In a polytomous logistic regression analysis of rs6983267 and rs1447295, assuming a multiplicative allelic risk model, the estimated genotype effects were comparable for aggressive and nonaggressive prostate cancer at diagnosis (**Supplementary Table 3**) ($P = 0.17$). We also investigated the possible effect of age on risk in an analysis of all

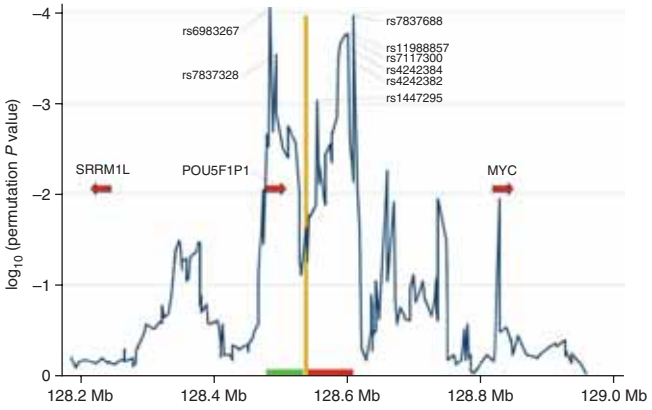

Figure 2 Association signal in the 8q24 region detected by ancestral recombination graph (ARG). Unphased genotypes of rs1447295 and 196 flanking SNPs were used to infer 100 ARGs[17], each one describing a possible mutation and recombination history for this region jointly for the controls, the nonaggressive cases and the aggressive cases of the PLCO cohort. An ARG gives a genealogy for every SNP position, and these genealogies were tested for association by placing putative causative mutations on the branches and performing a χ^2 test on a nine-cell contingency table including three phenotypes (aggressive and nonaggressive cases and controls) and three genotypes (4-d.f. χ^2 test)[17]. After combining this analysis across all genealogies at a position, we determined the significance by random permutation of the phenotypes (maximum number of permutations, 10^5). The log base10 of this evaluation, called $\log_{10}$(permutation *P* value), is plotted as a function of the SNP position along the chromosome. The eight locations that provided permutation *P* values $< 10^{-3}$ are indicated by rs numbers at this position. The vertical orange line indicates the position of the region with an estimated high recombination rate. The horizontal green and red lines indicate the position of the centromeric and telomeric haplotypes, respectively, that were reconstructed using the program PHASE and further studied for association using an ARG strategy (see main text and **Supplementary Note**). The positions of three notable genes are also indicated (horizontal red arrows).

prostate cancers in order to test for heterogeneity of genetic effects at ages above and below 65 years (**Supplementary Table 3**). The estimated genetic effects under a two-locus multiplicative allelic risk model did not differ significantly between age groups ($P = 0.18$).

Although the region of 8q24 analyzed in this report is frequently amplified in prostate tumors[20,21], it harbors few known or predicted genes. Furthermore, we did not observe an association between SNPs in the *MYC* gene (263 kb from rs1447295 in the telomeric direction) and prostate cancer risk.

Our results demonstrate how multiple SNPs within a chromosomal region in distinct blocks may be associated with disease risk. Although the rs6983267 G allele is associated with a lower relative risk than the rs1447295 A allele, it is substantially more frequent in populations of European ancestry (50% versus 11%). Based on our five studies, an estimate for the population attributable risk[22,23] (PAR) of prostate cancer associated with rs6983267 G is 21%, whereas the PAR for rs1447295 A is 9%; the estimated joint effects PAR for carriage of either or both is 27% (**Supplementary Table 4** online). The estimated joint and individual PARs suggest that the two loci substantially contribute to the population burden of prostate cancer. However, comparisons between PARs based on markers can be misleading when the correlation between the markers and the respective functional variants vary.

It is worth noting that subtle variations may occur within the European population[24]. In the Alpha-Tocopherol, Beta-Carotene study of Finns, the additional SNP in the telomeric block, rs7837688, that was tested across all studies showed a higher MAF and did not replicate, whereas the other studies showed a consistent, positive association (**Table 1**). This apparent discrepancy could be used to better define the region(s) harboring the functional variant(s) in the telomeric block similar to a previous approach to mapping *PRKCA* and multiple sclerosis[25,26].

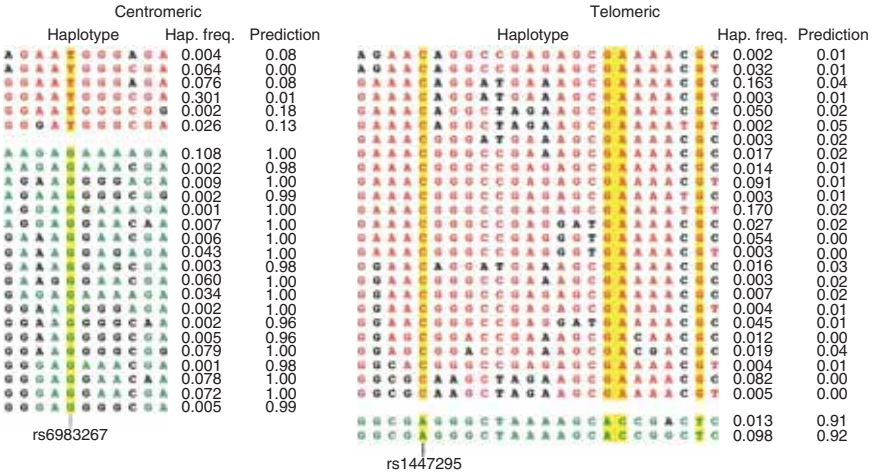

Figure 3 Centromeric and telomeric haplotypes of the 8q24 region associated with prostate cancer susceptibility in PLCO. The best pair of haplotypes for each individual was determined using PHASE[18,19] in two independent calculations for 20 and 27 SNPs located in either region flanking the recombination hotspot. For each region, the phased haplotypes were used to generate 100 genealogies[17], and the positions of the putative functional mutations were estimated for the eight locations indicated in **Figure 1** and **Figure 2**. The frequency with which each haplotype was predicted to carry the 'at-risk' mutation was then computed. Results are shown for rs6983276 (haplotypes with population frequencies larger than 0.001 defined by 11 contiguous SNPs: rs10956365, rs10505476, rs10808555, rs17467139, rs6983267, rs10505473, rs7837328, rs7014346, rs12375310, rs6995633, rs6999921) and rs7837688 (haplotypes with population frequencies larger than 0.002 defined by 24 contiguous SNPs: rs7830412, rs1447293, rs921146, rs4871799, rs1447295, rs9297758, rs13363309, rs11775749, rs16902169, rs13253127, rs6985504, rs16902173, rs12155672, rs1562432, rs1562431, rs4871808, rs4242382, rs4242384, rs7017300, rs11988857, rs9656816, rs7814251, rs7837688, rs6991990). Results from genealogies at position rs7837328 and rs1447295 were less contrasted. Those from positions rs4242382, rs4242384, rs7017300 and rs11988857 were identical to those of rs7837688, with the exception of the haplotype marked by an asterisk, predicted to carry an 'at-risk' mutation with a frequency of 0.36 for genealogies located at rs11988857. Positions in almost perfect linkage disequilibrium ($r^2 \sim 1$) with the imputed functional mutation are highlighted in yellow. 'Hap. freq.' = frequency of haplotype in all samples from PLCO; 'prediction' = frequency predicted to carry the 'at-risk' mutation.

Table 2 Odds ratios and 95% confidence intervals for the two-SNP unconstrained and multiplicative interaction models for all studies combined

			rs6983267		
			TT	GT	GG
rs1447295	CC	O	1.00 Referent	1.27 1.12–1.43	1.57 1.37–1.81
		M	1.00 Referent	1.24 1.17–1.32	1.55 1.37–1.75
	CA	O	1.46 1.17–1.84	1.76 1.49–2.08	2.25 1.85–2.74
		M	1.43 1.30–1.57	1.78 1.60–1.98	2.22 1.91–2.57
	AA	O	2.55 1.12–5.83	3.61 2.13–6.13	2.27 1.31–3.91
		M	2.05 1.70–2.46	2.55 2.10–3.09	3.17 2.55–3.94

Observed overall $P = 9.18 \times 10^{-22}$. Multiplicative overall $P = 1.69 \times 10^{-25}$. Analysis adjusted for age in 5-year intervals and study. O = observed (unconstrained); M = multiplicative model. See **Supplementary Table 3** for complete results of each study using unconstrained and multiplicative models.

Although it is possible that additional variants in the region of 8q24 could further modulate the risk of prostate cancer, the results of this GWAS did not identify other highly significant loci in men of European ancestry (**Fig. 1**). However, a previous admixture scan identified peaks in the region that could be important in men of other ancestral backgrounds[4]. We note that for the centromeric SNP, rs6983267, the 'at-risk' G allele has an estimated frequency of 0.98 in the Yoruban and 0.37 in the East Asian samples of HapMap compared with 0.50 in the controls of our combined studies (**Table 1**)[16]. These observations could partially explain the known ethnic disparities in prostate cancer incidence[2]. Further work is needed to identify common and uncommon variants across both regions (particularly in populations with different underlying genetic structures) to determine the optimal candidates for functional studies designed to confirm the causal variants in the 8q24 region.

METHODS

Detailed descriptions of the methods are provided in **Supplementary Methods**.

URLs. Cancer Genetic Markers of Susceptibility project: http://cgems.cancer.gov; HapMap: http://hapmap.org/; TagZilla: http://tagzilla.nci.nih.gov/; Aabel: http://www.gigawiz.com/

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

The HPFS study is supported by NIH grants CA CA55075 and 5U01CA098233-04. The ACS study is supported by U01 CA098710. The ATBC study is supported by NIH contracts N01-CN-45165, N01-RC-45035 and N01-RC-37004. F.R.S. is supported by an NRSA training grant (T32 CA 09001). P.F. is supported by a UK Engineering and Physical Sciences Research Council Grant (GR/S18786). M.M. is supported by the Wellcome Trust. N.O., R.B.H., S.W., K.Y., N.C., M.T., J.F.F., R.H., S.J.C. and G.T. are supported by the Intramural Research Program of the National Cancer Institute (US National Institutes of Health, Department of Health and Human Services).

COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Crawford, E.D. Epidemiology of prostate cancer. *Urology* **62**, 3–12 (2003).
2. Parkin, D.M. *et al. Cancer Incidence in Five Continents* (IARC Scientific Publications, Lyon, 2002).
3. Steinberg, G.D., Carter, B.S., Beaty, T.H., Childs, B. & Walsh, P.C. Family history and the risk of prostate cancer. *Prostate* **17**, 337–347 (1990).
4. Freedman, M.L. *et al.* Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc. Natl. Acad. Sci. USA* **103**, 14068–14073 (2006).
5. Amundadottir, L.T. *et al.* A common variant associated with prostate cancer in European and African populations. *Nat. Genet.* **38**, 652–658 (2006).
6. Schumacher, F.R. *et al.* Prostate cancer risk and 8q24. *Cancer Res.* (in the press).
7. Gohagan, J.K., Prorok, P.C., Hayes, R.B. & Kramer, B.S. The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial of the National Cancer Institute: history, organization, and status. *Control. Clin. Trials* **21**, 251S–272S (2000).
8. Prorok, P.C. *et al.* Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Control. Clin. Trials* **21**, 273S–309S (2000).
9. Phillips, M.S. *et al.* Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots. *Nat. Genet.* **33**, 382–387 (2003).
10. Calle, E.E. *et al.* The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. *Cancer* **94**, 2490–2501 (2002).
11. Chen, Y.C. *et al.* Sequence variants of Toll-like receptor 4 and susceptibility to prostate cancer. *Cancer Res.* **65**, 11771–11778 (2005).
12. Valeri, A. *et al.* Segregation analysis of prostate cancer in France: evidence for autosomal dominant inheritance and residual brother-brother dependence. *Ann. Hum. Genet.* **67**, 125–137 (2003).
13. The ATBC Cancer Prevention Study Group. The alpha-tocopherol, beta-carotene lung cancer prevention study: design, methods, participant characteristics and compliance. *Ann. Epidemiol.* **4**, 1–10 (1994).
14. Fearnhead, P. SequenceLDhot: detecting recombination hotspots. *Bioinformatics* **22**, 3061–3066 (2006).
15. Myers, S., Bottolo, L., Freeman, C., McVean, G. & Donnelly, P. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**, 321–324 (2005).
16. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
17. Minichiello, M.J. & Durbin, R. Mapping trait loci by use of inferred ancestral recombination graphs. *Am. J. Hum. Genet.* **79**, 910–922 (2006).
18. Stephens, M., Smith, N.J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**, 978–989 (2001).
19. Stephens, M. & Scheet, P. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *Am. J. Hum. Genet.* **76**, 449–462 (2005).
20. Nupponen, N.N., Kakkola, L., Koivisto, P. & Visakorpi, T. Genetic alterations in hormone-refractory recurrent prostate carcinomas. *Am. J. Pathol.* **153**, 141–148 (1998).
21. Cher, M.L. *et al.* Genetic alterations in untreated metastases and androgen-independent prostate cancer detected by comparative genomic hybridization and allelotyping. *Cancer Res.* **56**, 3091–3102 (1996).
22. Bruzzi, P., Green, S.B., Byar, D.P., Brinton, L.A. & Schairer, C. Estimating the population attributable risk for multiple risk factors using case-control data. *Am. J. Epidemiol.* **122**, 904–914 (1985).
23. Wacholder, S., Benichou, J., Heineman, E.F., Hartge, P. & Hoover, R.N. Attributable risk: advantages of a broad definition of exposure. *Am. J. Epidemiol.* **140**, 303–309 (1994).
24. Seldin, M.F. *et al.* European population substructure: clustering of northern and southern populations. *PLoS Genet.* **2**, e143 (2006).
25. Saarela, J. *et al.* PRKCA and multiple sclerosis: association in two independent populations. *PLoS Genet.* **2**, e42 (2006).
26. Willer, C.J. *et al.* Tag SNP selection for Finnish individuals based on the CEPH Utah HapMap database. *Genet. Epidemiol.* **30**, 180–190 (2006).